

7

Loi normale ou loi de Laplace-Gauss

- I. Définition de la loi normale
- II. Tables de la loi normale centrée réduite

S'il y avait une seule loi de probabilité à connaître, ce serait celle-là. Elle est importante en pratique car elle permet de représenter la variabilité de nombreux phénomènes naturels (la glycémie à jeun, le taux de division bactérienne, etc.). C'est elle, aussi, qui modélise les variations observées entre mesures successives d'une quantité, suite à l'erreur de mesure. Elle a une importance théorique considérable sur laquelle on reviendra dans la fiche 9 : un théorème (le théorème central limite) explique pourquoi on rencontre si fréquemment dans la nature des lois normales ; enfin elle est une clé nécessaire à la démarche statistique.

Elle fut découverte indépendamment par les mathématiciens Gauss en Allemagne (1809) et Laplace en France (1812).

I. Définition de la loi normale

A. Densité de probabilité de la loi normale

Définition : loi normale

Une variable aléatoire X suit une *loi normale*¹, ou loi de Laplace-Gauss ou loi de Gauss, si sa ddp s'écrit :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Elle est définie pour $-\infty < x < +\infty$.

Les deux paramètres μ et σ de la ddp sont respectivement la moyenne et l'écart type de X .

Plus que la formule (qui n'est pas utilisée en pratique), c'est la figure 7.1 qu'il faut étudier avec attention.

Bien noter que la ddp d'une variable aléatoire de moyenne μ et d'écart type σ est symétrique autour de μ et a deux points d'inflexion aux abscisses $\mu - \sigma$ et $\mu + \sigma$. Noter aussi qu'une loi normale est définie mathématiquement entre $-\infty$ et $+\infty$ alors qu'elle est utilisée en pratique pour représenter les variations interindividuelles de paramètres biologiques qui, évidemment, ne prennent que des valeurs finies (par exemple, la glycémie). La raison en est que la ddp de la loi normale sans être nulle

1. Le mot « normal » est employé couramment en biostatistique dans deux sens différents. Le bon sens permet « normalement » de se rendre compte quand on parle de glycémies « normales » (par opposition à pathologiques) et quand on parle de la glycémie distribuée selon une loi « normale » (dans le sens de distribuée selon une loi de Gauss).

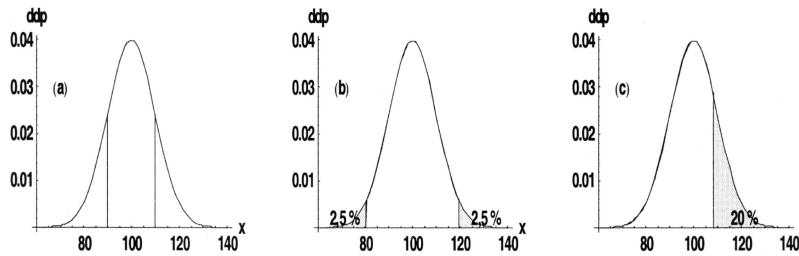


Fig. 7.1

- a. Exemple de ddp d'une loi normale : dans l'exemple $\mu = 100$ et $\sigma = 10$. Les verticales $\mu - \sigma$ et $\mu + \sigma$ sont représentées.
 - b. Il y a probabilité 5 % pour que la variable aléatoire X soit à l'extérieur de l'intervalle $\mu + u_{5\%} \sigma$ et $\mu - u_{5\%} \sigma$, ($u_{5\%} = 1,96$)
 - c. Il y a probabilité $\alpha_2 = 20\%$ pour que la variable aléatoire X soit supérieure à $\mu + z_{0,80} \sigma$ ($z_{0,80} = z_{80\%} = 0,84$).
- NB : Les valeurs sont données par les tables 3.1 à 3.4 ($u_{5\%} = 1,96$ dans la table 3.4 et $z_{80\%}$ dans la table 3.2).

« mathématiquement » prend des valeurs très faibles dès que l'on s'écarte suffisamment de μ : par exemple, une loi normale a seulement une chance sur un million de tomber au-delà de 5 écarts types de part et d'autre de la moyenne.

Quelques valeurs particulières *doivent* être mémorisées tant elles sont sans cesse utilisées.

Résultat : valeurs limites importantes dans la loi normale de moyenne μ et d'écart type σ

- il y a 10 chances sur 100 pour que $X < \mu - 1,65 \sigma$ ou $X > \mu + 1,65 \sigma$;
- il y a 5 chances sur 100 pour que $X < \mu - 1,96 \sigma$ ou $X > \mu + 1,96 \sigma$;
- il y a 1 chance sur 100 pour que $X < \mu - 2,58 \sigma$ ou $X > \mu + 2,58 \sigma$;
- il y a 1 chance sur 1 000 pour que $X < \mu - 3,30 \sigma$ ou $X > \mu + 3,30 \sigma$.

Dans beaucoup de problèmes statistiques, on se servira en particulier de la deuxième ligne : une variable aléatoire X distribuée normalement a 5 chances sur 100 de présenter un écart à la moyenne supérieur à 2σ (on arrondit en général 1,96 à 2). Dit autrement, 95 % des sujets sont distribués dans une étendue de 4σ . Par exemple, si l'on nous dit que, dans une certaine population, chez le sujet adulte « normal » (= non diabétique), la glycémie est distribuée selon une loi normale de moyenne 4,8 mmol/L et d'écart type 0,4 mmol/L, on déduit immédiatement que 95 % des sujets « normaux » de cette population ont une glycémie comprise entre 4,0 et 5,6 mmol/L.

B. Loi normale centrée réduite

Définition : loi centrée réduite

On appelle loi normale centrée réduite la loi normale de moyenne 0 et de variance 1.

Une variable suivant la loi normale centrée réduite est notée Z .

Si X est de moyenne μ et d'écart type σ , $Z = \frac{X - \mu}{\sigma}$ suit une loi normale centrée réduite.

Pour montrer que Z est de moyenne 0 et de variance 1, il suffit d'appliquer le résultat donné pour le changement de variable linéaire (intuitif, cf. aussi fiche 8), en posant $a = 1/\sigma$ et $b = -\mu/\sigma$.

7 Bases de calcul des probabilités

Ce changement de variable est en pratique *très* utile : des tables de la f.r et de la ddp de Z ont été construites et peuvent être utilisées à propos d'une variable X quelconque en utilisant ce changement de variable (ou l'inverse $X = \mu + \sigma Z$).

C. Addition de variables normales indépendantes

Bien que la définition précise de l'indépendance de variables aléatoires soit traitée dans la fiche 8, on donne tout de suite ce résultat.

Résultat

Si U et V sont deux variables aléatoires *indépendantes* distribuées normalement, toute combinaison linéaire des deux de la forme $aU + bV$ est distribuée normalement.

Sa moyenne vaut $a\mu_U + b\mu_V$.

Sa variance vaut $a^2\sigma_U^2 + b^2\sigma_V^2$.

II. Tables de la loi normale centrée réduite

Quatre tables sont fournies en annexe p. 205 et suivantes. Deux sont destinées à relier la probabilité que Z soit inférieure à une valeur donnée z avec cette valeur z (c'est donc la f.r. de Z qui est ainsi tabulée) ; les deux autres, très utiles en statistique, donnent la probabilité que $|Z|$ (la valeur absolue de Z) soit supérieure à une valeur donnée en fonction de cette valeur. Le tableau 7.1 représente les objectifs de ces quatre tables. Compte tenu de leur importance pratique, il est indispensable d'effectuer un grand nombre d'exercices d'utilisation de ces tables.

Tableau 7.1. Objectifs d'utilisation des 4 tables concernant la loi normale

Table 3-1	Table 3-2	Table 3-3	Table 3-4
On donne une valeur z . On se demande quelle probabilité p a Z d'être inférieure à z .	On donne une probabilité p . On se demande quelle est la valeur de z pour laquelle il y a une probabilité p que Z lui soit inférieure.	On donne une valeur de u . On se demande quelle probabilité Z a d'être au-dessous de $-u$ ou au-dessus de $+u$.	On donne une probabilité p . On se demande quelle est la valeur u telle qu'il y ait probabilité p d'être en dessous de u ou au-dessus de u .
<i>Exemple</i> : 98,61 chances sur 100 pour que $Z < 2,2$.	<i>Exemple</i> : la valeur $z = 1,645$ a 5 chances sur 100 d'être dépassée.	<i>Exemple</i> : Il y a une probabilité de 10,96 % pour que $ Z > 1,60$.	<i>Exemple</i> : Il y a 98 chances sur 100 pour que Z soit entre $-0,025$ et $+0,025$.

Les tables 3.3 et 3.4 sont appelées « tables aux extrémités », car elles permettent de calculer les probabilités à l'extérieur d'un intervalle encadrant la moyenne (en anglais : *two tails tables*). Les tables sont données toutes les 4 pour faciliter les calculs : éviter quelques inter-

polations et quelques divisions par deux ; mais une seule suffirait. C'est un exercice simple que de déduire, à partir d'une de ces tables, les 3 autres.

Il est facile de déduire (fig. 7.2), par exemple, la table 3.1 à partir de la table 3.3 : il y a 5 chances sur 100 d'après la table 3.3 pour que $|Z| > 1,96$. Il y a donc 2,5 chances sur 100 pour que $Z < -1,96$ et 2,5 chances sur 100 pour que $Z > 1,96$; en effet, on lit dans la table 3.1 à la ligne $-1,96$ la probabilité 2,5 % et à la ligne 1,96 la probabilité 97,5 %.

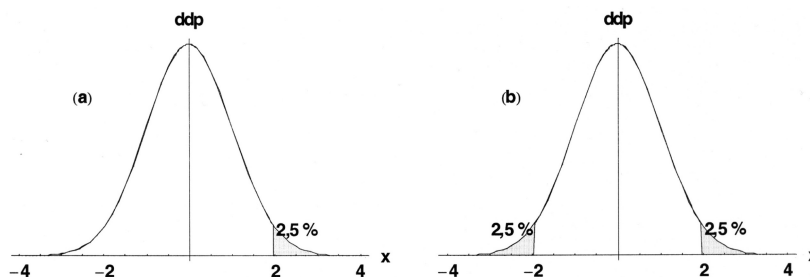


Fig. 7.2

Dans la loi normale réduite centrée, on a $Pr(Z > 1,96) = 2,5 \%$ (figure de gauche) et $Pr(|Z| > 1,96) = 5 \%$ (figure de droite). On a $z_{2,5\%} = 1,96$ et $u_{5\%} = 1,96$.

Notation

u_α désigne la valeur qu'une loi normale réduite centrée Z a probabilité α de dépasser en valeur absolue (tables aux extrémités 3.3 et 3.4) : $Pr(|Z| > u_\alpha) = \alpha$.
 z_p désigne la valeur qu'une loi normale réduite Z a probabilité p de dépasser ($Pr(Z > z_p) = p$).

■ u_α est aussi noté, dans beaucoup de livres francophones, ϵ_α

Puisque la loi normale est symétrique, il vient immédiatement que $u_\alpha = z_{\alpha/2}$. Le lecteur le vérifiera sur les tables 3.1 à 3.4 (p. 208 à 212). La relation entre p et z_p s'obtient directement à partir des tables 3.1 et 3.2 en lisant à la ligne $1 - p$, puisque ces tables donnent les probabilités pour $Z < z$ et qu'ici, on cherche $Pr(Z > z)$.

Exercice 7.1

On suppose que les valeurs d'un dosage sont distribuées selon une loi de Gauss ; dans une population, 70 % des sujets ont une valeur de dosage supérieure à 120 et 10 % ont une valeur supérieure à 180. Calculer μ et σ .

Réponse

On utilise cette fois la relation $X = \mu + \sigma Z$.
 La valeur z dépassée par 70 % des sujets (cf. table 3.1) est $-0,524$. Celle dépassée par 10 % des sujets est $1,282$. On a donc les deux équations à deux inconnues :
 $120 = \mu - 0,524 \sigma$ et $180 = \mu + 1,282 \sigma$
 D'où l'on tire $\sigma = 33,3$ et $\mu = 137,4$.

7 Bases de calcul des probabilités

Exercice 7.2

Si dans une population 50 % des sujets de sexe masculin ont un poids supérieur à 70 kg et si l'écart type de la distribution est 10, quel % des sujets pèse plus de 100 kg ? Quelle hypothèse est nécessaire pour le calcul ?

Réponse

$\mu = 70$ kg (car la moyenne est égale à la médiane dans une loi normale) et $\sigma = 10$. On effectue la transformation habituelle $Z = \frac{X - 70}{10}$. À $x = 100$, correspond $z = 3$. Grâce à la table 3.1, on trouve la probabilité recherchée qui est de $1 - 0,99865 = 0,00135 = 0,135$ %.
Le calcul fait l'hypothèse que le poids est distribué normalement.

Exercice 7.3

Si la glycémie à jeun est distribuée normalement dans une certaine population chez les sujets (hommes, adultes) non diabétiques avec une moyenne de 5,5 mmol/L et un écart type de 0,2 mmol/L et chez les sujets dont le diabète a été découvert récemment avec une moyenne de 6,0 mmol/L et un écart type de 0,3 mmol/L, quel % des sujets normaux et quel % des sujets diabétiques ont une glycémie supérieure à 6 mmol/L ?

Réponse

Chez les non-diabétiques $z_{nd} = \frac{x - 5,5}{0,2}$ et chez les diabétiques $z_d = \frac{x - 6}{0,3}$.
Quand x vaut 6, z_{nd} vaut 2,5 et la probabilité cherchée (table 3.1) vaut $1 - 0,99379 = 0,00621$. Comme 6 est la moyenne chez les diabétiques, la probabilité, chez eux, de dépasser 6, est 50 %.

Exercice 7.4

Dans une population, on tire au sort 10 000 sujets et on dose la quantité X que l'on suppose distribuée normalement. On range les 10 000 valeurs en ordre croissant. La 251^e vaut 72 et la 9 750^e vaut 116. Évaluer μ et σ .

Réponse

95 % des sujets sont compris entre $\mu - 2\sigma$ et $\mu + 2\sigma$. Ici, 95 % des sujets sont compris entre 72 et 116. On en tire les évaluations suivantes : $\mu = 94$ et $\sigma = 11$ (« évaluations » car, sur un autre échantillon de 10 000 sujets, la 251^e valeur n'aurait certainement pas été à nouveau exactement 72...).

Exercice 7.5

On peut calculer la probabilité pour qu'une variable distribuée selon une loi de Poisson de paramètre suffisamment grand (supérieur à 30) tombe dans un intervalle de valeurs en approximant cette loi par une loi normale. On suppose que le paramètre vaut 8 (très en dessous de 30, donc). Calculer la probabilité pour que la variable soit en dessous de 3, et comparer avec le calcul exact (on donne $e^{-8} = 3,3546 \times 10^{-4}$).

Réponse

Par l'approximation normale, on calcule la probabilité pour qu'une loi normale X de moyenne 8 et de variance 8 soit supérieure à 16. Pour

cela, on forme $Z = \frac{X - 8}{\sqrt{8}}$ et on cherche dans la table 3.1 la probabilité

correspondant à $Z = \frac{3 - 8}{\sqrt{8}} = -1,77$ qui est environ 4 %.

La probabilité exacte s'obtient en sommant les probabilités d'obtenir 0, 1, 2, et 3 dans une loi de Poisson. On trouve 4,24 %, ce qui est proche de 4 %.

POINTS CLÉS

- ▶ 95 % des valeurs d'une variable distribuée selon une loi normale sont comprises entre la moyenne $- 2$ écarts types et la moyenne $+ 2$ écarts types.
- ▶ Lorsqu'on dispose d'un échantillon, si on ôte 5 % (approximativement) de ces valeurs extrêmes, l'étendue séparant la plus petite valeur de la plus grande de cet échantillon tronqué permet de trouver l'ordre de grandeur de 4 écarts types.